Journal of Science
University of Ha'il

University of Ha'il–Journal of Science (UOHJS)

Volume IV - Second Issue

ISSN: 1658-8096
e-ISSN: 1658-8800

December 2023

Developing an Intelligent Model for Assessing Graduates' Skills Using Dual Kernel Support Vector Machine

Maher Mohammed Zakariya*, Fahad Mazaed Alotaibi*

Faculty of Computing and Information Technology (FCIT), King Abdul Aziz University (KAU), Jeddah, Saudi Arabia

* *Corresponding authors; E-mail: mzakariyafazal@stu.kau.edu.sa, fmmalotaibi@kau.edu.sa*

Abstract:
In this article, the effectiveness of higher education institutions is evaluated based on the performance of their graduates. Therefore, the primary aim of these institutions is to provide top-quality education to their students. Evaluating the skills of graduates is critical in assessing the educational system's effectiveness. This study's objective is to develop an intelligent model that utilizes a Dual Kernel Support Vector Machine (DK-SVM) to assess graduates' skills. The proposed DK-SVM model incorporates two kernel functions, namely Linear and Radial Basis Functions, to enhance the assessment process's efficiency and accuracy. The Dual Kernel SVM (DKSVM) model, can reliably and effectively evaluate the skills of graduates based on the extracted features. Using the questionnaire data obtained from King Abdulaziz University, the researchers developed this model to rate job performance. The experiment's results were evaluated utilizing various performance metrics, such as Accuracy, Recall, Precision, and f1-score. The proposed method attains an accuracy of 92%, recall of 90%, precision of 91%, and f1-score of 89%. The experimental results demonstrate that the proposed DK-SVM model outperformed other methods such as Support Vector Machine (SVM), Multiple Linear Regression (MLR), and Multi-layer Adaptive Neuro-Fuzzy Inference System (MANFIS).

Key words: *Data Mining, Predictive Models, Graduate Skill Set, Dual Kernel Support Vector Machine, and Early Prediction, Machine Learning.*

1. INTRODUCTION

Graduate employment is an increasingly important topic in higher education, as graduates' career opportunities are a key factor in assessing a university's ranking (McMurray et al. 2016, James et al. 2013). It is crucial to evaluate graduates' skills and their readiness for workplace development and requirements. The main goal of this study is to measure graduates' employability skills and to examine how these skills are perceived in higher education. With a large number of graduates competing in the job market, employers face challenges in finding suitable candidates for available positions. Employers look for the most qualified employees, selecting candidates based on their skills and qualifications

(Alam et al. 2021, Mengash, 2020). In addition to specific skills, employers also expect workers to have a strong work ethic, discipline, and the ability to adapt to various tasks and roles. Discussions about graduate employment often focus on the role of business graduates and the responsibilities of universities in preparing them for the workforce (Alshanqiti and Namoun 2020, Abdelmoumin et al. 2021). To meet the demands of diverse employment opportunities, traditional assessment methods, such as direct interviews, syllabus reviews, and reference checks, are utilized to evaluate graduates' skills. However, the influence of these factors varies across different academic programs and cohorts of graduates (Chui e al 2020, Alhazmi and Sheneamer 2023). Despite their usefulness, universities face challenges in analyzing large academic datasets to predict graduate performance, as this process can be time-consuming, biased, and subjective (Chui et al. 2020).

Detecting the performance of graduates provides several benefits, such as selecting suitable courses, identifying struggling students who may require additional support to complete their coursework, and predicting which students may quit the course due to difficulty, which can affect graduate retention rates (Waheed et al. 2020). Accurately predicting graduate achievement is a challenging task, and it requires the consideration of various factors and situations. To address this challenge, new intelligent methods such as machine learning are being developed to evaluate the performance of graduates efficiently and accurately. It is important to understand enhancing the achievements of the graduate, decrease dropout risks, and devise suitable programs for preventing factors. The dual kernel SVM method is a useful tool that allows for the clustering of related factors, which can help in predicting future course grades. Therefore, we propose a smart model that utilizes the Dual Kernel SVM algorithm to assess the skills of graduates.

In graduating the skills the predictive model is designed to perform the classification of numerical regression surface. The mismatch among the skills of graduates shows the impacts of employers and institutions. This is performed by machine learning approaches and performs classification by a data mining algorithm that is designed by a group of knowledge-based data. The employability signals of the graduates are predicted by a machine learning algorithm to improve employability as early as possible. In this process, it is significant to generate the appropriate knowledge and skills by assessing initial stage studies. The main contribution of this paper is summarized as follows:

- Applying data preprocessing techniques in the input dataset to obtain improved results.
- This paper presents a model to assess the skill set of the graduates and classify them into low, high, and satisfactory.
- Using the King Abdulaziz University questionnaire dataset, the proposed model is evaluated in terms of recall, precision, F1 score, and accuracy.
- The rest of this paper is organized as follows. Section II contains the literature review. Section III presents the proposed method, Section IV the Experimental Results Analysis At last, the conclusion is provided in section V.

2. LITERATURE SURVEY

Meegahapola and Thilakarathne (2019) developed an interactive learning tool architecture to enhance the fundamental knowledge of students and educate them on recent developments. Two learning tools were utilized, the first tool to improve student knowledge, and the second tool to improve their understanding of smart renewable power networks. However, navigation problems were identified in the tools.

Rivas et al. (2019) presented a neural network to enhance academic performance by detecting patterns of behavior using collected data from student groups. The model's performance was evaluated using metrics like recall, precision, and f1-score, but it required a long time for training. Almufarreh et al. (2023) developed the Quality Teaching and Evaluation Framework (QTEF) to ensure teachers' performance. Simple Linear Regression (LR) and Multiple Linear Regression (MLR) were utilized for forecasting and detecting. The model's performance was evaluated through experimental results, but it required more data.

Ulloa-Cazarez et al. (2021) introduced the Multi-layer Adaptive Neuro-Fuzzy Inference System (MANFIS) to predict the performance of students online. The process of training and testing was done with the use of the online education dataset. However, the computational cost of this model was high. Mengash (2020) presented a prediction model that utilized a linear regression approach for predicting student performance. The model utilized data mining methods such as Naive Bayes, Support Vector Machine (SVM), Decision Tree, Artificial Neural Network (ANN), and. The data were collected from Princess Nourah bint Abdulrahman University (PNU). However, ANN requires high computational power.

3. PROPOSED METHODOLOGY

The proposed model's architecture is illustrated in Figure 1. The model first obtains data from King Abdulaziz University datasets, including scores and gender. The collected data undergo pre-processing, followed by feature selection and user set classification using the DK-SVM method. A multiple kernel function is employed to process the input features. The following sections provide a detailed explanation of the proposed model's execution. The steps in the proposed model are as follows:

- Data Collection Module: This module is responsible for gathering relevant information from multiple sources such as social media profiles, academic records, and job applications. The data was collected from King Abdulaziz University by conducting interviews with English graduate students. Out of 469 candidates (total sample numbers), the majority of them were male with the age range of 23 to 29 and above 29 (considered as total attributes). From the total feature samples, 281 are considered to assess the skills of graduated candidates. The majority of the candidates completed their bachelor's and few did their masters with working experience of one to three years. Only very few candidates are familiar with the recruitment as well as assessment process for hiring graduates. Except for very few candidates, the remaining candidates agree with both the limitations as well as weaknesses in the current assessment methods. Also, only a few agreed with the efficiency and effectiveness of current assessment models. The majority of the candidates strongly agreed with the potential impacts and consequences of hiring candidates with inadequate skills for specific job roles. 90% of the candidates agreed to prefer smart models to assess the skills of various graduates before hiring them. Almost 50% of the candidates accepted the potential barriers as well as limitations when using intelligent models for assessing the skills of graduates prior to their employment. Thus the data was collected based on the individual feedback provided by each graduate. The collected information is stored in a centralized database for further processing.

Some of the features extracted by graduate students are personal qualities, academic performance, and work experience.

Table 1 depicts the qualities of graduate students with a few questionnaires to determine various skills based on their feedback.

Table 1. Questionnaires regarding experience validation of graduated candidates

Number	Questions	Number	Questions	Features
Q1	Why the gender equality is significant for education?	Q2	Which category age group is best for studying?	Personal qualities
Q3	Do you have any experience in any jobs?	Q4	Why Education is the most important in our life?	Experience
Q5	In which field are more familiar for hiring graduates?	Q6	Are you satisfied with the efficiency and effectiveness of the current assessment?	
Q7	What is the significant process while recruiting the graduate?	Q8	What is the consequence and impacts obtained in hiring the candidate for specific roles?	
Q9	What needs does the organization need to assess skills?	Q10	Whether any weakness determined in the assessment process?	Academic performance
Q11	Which is more significant for the organization?	Q12	What are the limitations determined to analyze the skills of graduates before they are employed?	
Q13	What are the skills needed to enhance the organization based on smart techniques?	Q14	Which academic, personal management, and teamwork skills are required to improve the organization?	
Q15	What are the difficulties in hiring graduates with insufficient skills?	Q16	How to improve the efficiency of skill assessment?	Job qualities
Q17	Why do various companies prefer a smart model?	Q18	Which soft skill is ranked more significant?	
Q19	Differentiate the soft skills individually?	Q20	Does the perception differ based on the environment?	
Q21	Which employability skill is required to perform an individual job role?			

- Data Preprocessing Module: This module is used to clean and preprocess the gathered information to eliminate errors and ensure that the information is in an appropriate format for machine learning. The data is cleaned, normalized, and transformed to prepare it for feature selection and training.

- Feature Selection Module: This module selects the most relevant features to train the model. It identifies factors that affect graduates' abilities, such as work experience, academic performance, and other relevant skills.
- Dual KSVM Module: This module trains the model using the selected features and preprocessed data. The machine learning algorithm is chosen according to the available data and the nature of the problem.

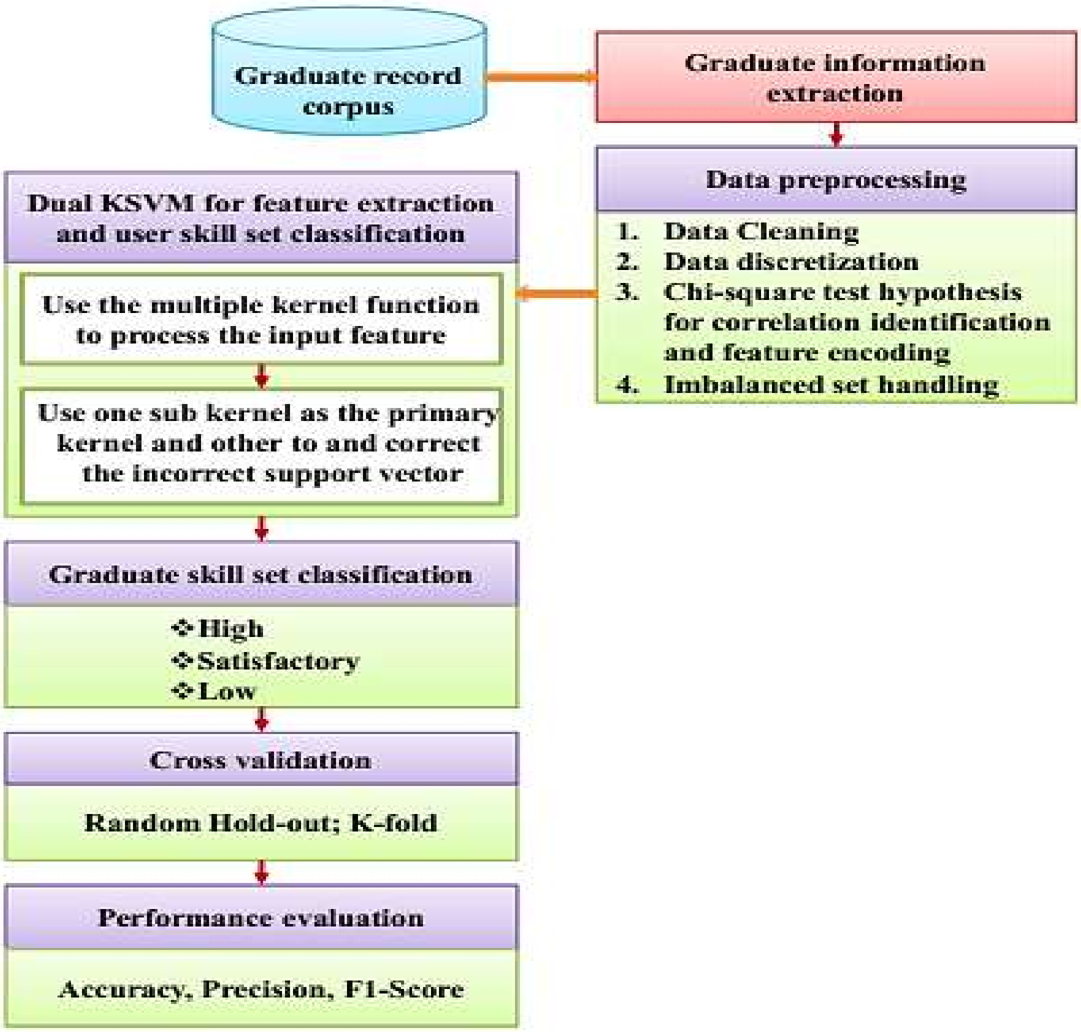

Figure 1. Architecture of proposed model.

3.1. System Model

The proposed system aims to assess the skills of graduates using a Dual Kernel Support Vector Machine (DK-SVM) model. The collected data is preprocessed, and using DK-SVM the feature selection and user set classification are carried out. Multiple kernel functions are used for processing the input features. The following represents the system model of the proposed system

$$z = g(i) \tag{1}$$

Here, i denotes a collection of features that indicate the graduates' skills such as personal qualities, academic performance, and work experience, and z denotes the predicted job performance rating based on graduates' assessed skills. $g()$ represents a function that uses a Dual Kernel Support Vector Machine (SVM) to transfer the input variable i to the output variable z . To increase the model's efficiency and accuracy the Dual Kernel SVM function combines the linear and radial basis functions as two kernel functions. The radial basis function is used to capture nonlinear relationships whereas the kernel function is used to capture linear relationships within the output variable and the input features. To learn the relationship between the input features and the output variable, using a dataset of graduates' skills, the model is trained, and the assessments of their performance on the job correlate to those skills. Based on their evaluated skills, new graduates' job performance ratings are then predicted using the trained model. The foremost step in generating a smart model is to describe the issue you wish to rectify. In this topic, the problem is to consider the graduate's skills according to particular criteria, such as work experience, academic performance, and further suitable elements.

3.2. Data processing

The first step in the proposed model is to gather comprehensive, accurate, and relevant data from different sources such as job applications, resumes, social media profiles, and academic records. This data is necessary to ensure that the model is based on valid information. The goal is to analyze the important factors in the performance of graduates and submit the features and students at the end stage, to detect the performances of the graduates who registered the data earlier.

3.2.1. Data preprocessing

Once the data is collected, preprocessing will take place to terminate any errors or inconsistencies. This model assists in data transformation, data normalization, and data cleaning to ensure that the data is in a suitable format for machine learning techniques.

Data transformation: Data transformation is the process of structuring and converting the data into a usable format that can be analyzed to assist in the decision-making process thereby propelling the organization's growth.

Data Normalization: Data normalization is the process of organizing various data in a particular database that includes creating and establishing relationships among tables thereby protecting the data and removing the consistencies as well as data dependency.

Data Cleaning: Data cleaning is the process of eliminating and fixing incorrect, incomplete as well as duplicate data within the dataset.

3.2.2. Data discretization

In this study, a discretization mechanism is employed to transform the graduates' numerical values grades to nominal values that express the classification problem by the given dataset. Here, an equal-width binning technique is involved in this phase based on the university standard grade as illustrated in Table 2. To make it easier for the algorithm to learn and achieve better results, Depending on the graduate's grades the input features were categorized: Very Good, Good, Outstanding, Fail, and Poor. Based on grades, the two nominal intervals pass and fail were divided. Furthermore, the relationship between the categorical variables based on the graduates' performance was identified using the chi-square test as a statistical measure.

Table 2. Discretization Step Results.

Input features	Marks	Class label (Skillset)
Outstanding	85 - 100	High
Very good	70 – 84	
Good	50 - 69	Satisfactory
Poor	35 - 49	Low
Fail	< 35	

In order to compute chi-square statistic, the following steps are to be followed:

1. A contingency table containing observed frequencies for each category is developed.
2. Compute the expected frequency for every category under the null hypothesis.
3. The chi-square test hypothesis has zero connection among categorical variables. This testing was carried out depending on a 0.05 significance level.
4. If the hypothesis is refused, the statistical correlation will take place among categorical variables.
5. The chi-square equation (2) is proposed in order to recognize the connection among the categorical variables, as shown.

$$C^2 = \sum \frac{(l-v)^2}{v} \tag{2}$$

From equation (2), v be the expected frequency and l be the observed frequency. Upon conducting a chi-square test, at the start of the academic year, a significant correlation was detected between the grades of the students and the data structures course in the following year.

3.2.3. Feature Encoding

The input features are characterized in feature encoding as classified data but machine learning algorithms can't process them straightly. Therefore, it is essential to change the categorical data to a numerical format that the algorithms can handle. For this purpose, label encoding is used, which assigns a unique integer to each nominal variable. In this study, the input features and output labels are transformed into integers. Specifically, the categories "Fail," "Poor," "Good," "Very Good," and "Excellent" are mapped to the integers 4, 3, 2, 1, and 0, respectively. The output label "Fail" is mapped to 1, and "Pass" is mapped to 0.

3.2.4. Oversampling

For solving the imbalanced dataset problems and expanding the number of classes from the dataset this process generates a unique sample and this method is appropriate for data that is small in size. In this study, several sample oversampling approaches were employed including Adaptive Synthetic Sampling (ADASYN), Synthetic Minority Oversampling Technique (SMOTE), and Random Oversampling (ROS).

3.2.5. Feature Extraction

The feature extraction includes relevant features extracted from the data and converted into a format that uses the Dual Kernel SVM model. The following equation represents this step's equation model

$$I = [i1, i2, i3, \ldots\ldots\ldots, im] \tag{3}$$

Here, I is the feature matrix that denotes the extracted features for each graduate. $i1, i2, i3, \ldots\ldots\ldots, im$ are the individual features that are extracted from the data, such as personal qualities, academic performance, and work experience. The following equation represents each feature is:

$$i_s = g_s(e_s) \tag{4}$$

The above equation, e_s denotes the s^{th} feature raw data, such as graduates' work experience or academic performance. g_s indicates the feature transformation function that maps the raw data e_s to a feature space suitable for the Dual Kernel SVM model. Depending on the type of data being transformed a different feature transformation function g_s can be used. A more sophisticated transformation function for work experience would involve calculating the number of years of experience and weighing it depending on the importance of the job responsibilities, while a basic transformation function for academic performance might be to normalize the grades to a scale of 0 to 1. The Dual Kernel SVM (DKSVM) model, can reliably and effectively evaluate the skills of graduates based on the extracted features. The feature matrix I can be used as the input.

3.3. Proposed DKSVM for graduate skill set analysis

A dynamic dual-kernel SVM is proposed to be employed in detecting the skills and course scores of various graduates. (Qiu et al. 2020). Although one dedicated kernel is utilized in SVM, DK-SViM utilizes multiple kernel functions to obtain the input feature with higher accuracy. Consider $A = \{B_r^a, ca\}$ as the input vectors, here B_r^a and ca denote the a^{th} samples and the label B_r^a respectively. The dual kernel is represented in the following equation:

$$L(B_r^a, B_r^d) = \sum_{r=1}^{4} \sigma_r L_{DUAL}^r (B_r^a, B_r^d) \tag{5}$$

$$\sigma_r L_{DUAL}^r = (\sigma_r - \Psi) L_{f_1}^r + \Psi L_{f_2}^r \tag{6}$$

Here, L_{DUAL}^r is the dual-kernel and the sub-kernels are represented as L_{f_1} and L_{f_2}. The sub-kernel $L_{f_1}^r$ is utilized as the primary kernel and $L_{f_2}^r$ is utilized in order to correct the miscategorized support vectors $L_{f_1}^r$. The coefficient of sub-kernels is denoted as $\Psi \subset [0,0.1]$. σ_r is the coefficient of several kernels and it fulfills $\sum_{r=1}^{4} \sigma_r = 1$. It is also notable that two features of A resemble one another if they belong to the same graduate skill class. As a result, the KSVM prediction can include label information from the training samples. Here it is also proposed to utilize the label information that will change with the KSVM. The following equation expresses the dynamic dual-kernel for the training process.

$$L_f(B_r^a, B_r^d) = \exp(\log L(B_r^a, B_r^d) + M) \tag{7}$$

$$M(B_r^a, B_r^d) = \begin{cases} \alpha & c_a = c_d \\ 0 & c_a \neq c_d \end{cases} \tag{8}$$

Here, the dynamic factor α denotes the weight information of the label. The dynamic ideal kernel is represented by $M(B_r^a, B_r^d)$. The following equation shows the new synchrophasor samples of the DKSVM are calculated by the Neumann divergence.

$$L_f(B_r^v, B_r^u) = -L(B_r^v, B_r^u) + L(B_r^v, B_r^u) NL(B_r^a, B_r^d) \tag{9}$$

Here, $N = L^{-1}\left(L_f\left(B_r^a, B_r^d\right) + L\right)L^{-1}$. The prediction samples are denoted by $B_r^v, B_r^u \subset A$. Finally, the graduates' skill class decision function can be expressed as:

$$\bar{c} = sign\left(P^M L_f\left(B_r^v, B_r^u\right) + h\right) \tag{10}$$

Here, h and P^M represent the bias term and learned weight vector respectively. The research problem was formalized and translated into the proposed solution. The structure of the dataset was analyzed and organized based on the labeled classes and feature types. The data was obtained from an educational system that records graduate data chronologically, usually by academic semesters. The dataset was split into two parts based on time, current and past semesters, to develop the models. For the semesters the schedule data is utilized in terms of detection (Alshanqiti and Namoun 2020). The input $C^{p\times o}$ defines features of the student course, where p is the input instance numbers, and o is the feature number, such that $\{p,o\} \subset \mathrm{M}$. The preferred results contain the performance of the graduate, described as the entire assessment grade R^P, i.e. $r_u = [0,100]$, and a group of factors $RE^{p\times l}$, such that $RE_{p,q} = \{re_{p,q} : q \subset 1,2,...,0\}, l \subset \mathrm{M}$, and each target $re_{p,q}$ should be above a quantitative value. Here the course and graduate features were differentiated $\ddot{Y}$ and $\dot{Y}$, individually.

3.4. Computational complexity

The complexity of the DKSVM algorithm is determined by the standard SVM solvers that are employed to generate an optimal solution by generating the matrix ω. The training complexity of the SVM is determined by $O\left((n+m)^3\right)$ and this operation is applied for each iteration. In the presence of high datasets, the gradient process creates complications and makes the process impractical. But the computational complexity data diminished the burden and validated the scheme to attain optimal solution. The optimization issues are solved by representing the matrix $f(y) = v_N \varphi(y)$ and this addresses the scalability issues in SVM learning. The testing complexity is determined from the training phase and terminated the entire process by predicting the samples from the target domain.

4. EXPERIMENTAL RESULTS AND ANALYSIS

The proposed model is evaluated utilizing the questionnaires and interviews conducted using the graduate students from King Abdulaziz University who selected English as their foreign language. These details were then stored in the admission and registration Deanship office and we obtained the required ethical approval from the King Abdulaziz University for data access. The interview conducted demonstrates the significance of individual feedback in evaluating their graduate skills. The details are then saved in the spreadsheet format using Microsoft Excel and then converted to the standard format using the Waikato Environment for Knowledge Analysis (WEKA) tool. During dataset preprocessing, we eliminated the student's personal information which includes their name, address, and ID along with the duplicate removal. For cross-validation, two general techniques were used in this article. The two techniques are elaborated on below.

- Random Hold-Out Method: The chosen dataset is split into two parts, in the training process, 80% of data is employed and 20% of data is employed in the testing process in this technique.
- K-Fold Cross-validation: K-Fold Cross-validation is utilized for improving the performance of the machine-learning approaches. For testing and training, it uses the whole dataset. If the dataset is tiny, this technique is an effective tool as described in this article. The dataset is split into two equal sizes of K subsets. For testing one of the subsets is used and the other subset is used for training. The output is acquired by taking the mean output of the testing set. The stratified K-fold method is used in this article, It is a type of K-fold cross-validation and it is employed with separation issues. So a stratified 5-fold cross-validation is utilized.

The machine learning methods used in this article are Dual Kernel Support Vector Machine (DK-SVM), and Support Vector Machine (SVM), Multiple Linear Regression (MLR), Multi-layer Adaptive Neuro-Fuzzy Inference System (MANFIS).

4.1. Performance Metrics

Some performance metrics like Precision, Accuracy, Recall, and f1-score are utilized in order to evaluate the performance of the model. This module evaluates the performance of the trained model using validation data. The metrics are compared against other models to identify the best one. The mathematical representation of these performance metrics is as follows:

- Accuracy: It is mostly utilized as a performance parameter. Accuracy is indicated as (%) and it is the percentage of the accurate results that are predicted.

$$Accuracy = (T_P + T_N)/(T_P + T_N + F_P + F_N) \tag{11}$$

T_p indicated as the true positive rate, F_P represented as the false positive rate, T_N denoted as the True negative rate, F_N indicated as the False Negative Rate.

- Recall: It is the percentage of positives accurately predicted as positives.

$$\mathrm{Re}\,call = T_P /(T_P + F_N) \tag{12}$$

- Precision: It calculates the positive classes and it is the percentage of the accurate positive observation.

$$\Pr ecision = T_P / (T_P + F_P) \tag{13}$$

- f1-score: It expresses the balance between recall and precision. f1-score also known as harmonic mean and requires false negatives and false positives

$$f_1 - score = 2 \times \operatorname{Re} call \times \Pr ecision, / (\operatorname{Re} call + \Pr ecision) \tag{14}$$

4.2. Performances analysis

Figure 2 shows the performance of the various techniques used, based on accuracy, recall, precision, and F1-measure. By comparing the proposed DK-SVM method with the other methods, the proposed method shows better performance. It outperformed the other methods when compared with other techniques such as SVM (Mengash 2020), MLR (Almufarreh, et al. 2023), and MANFIS (Ulloa-Cazarezet al. 2021).

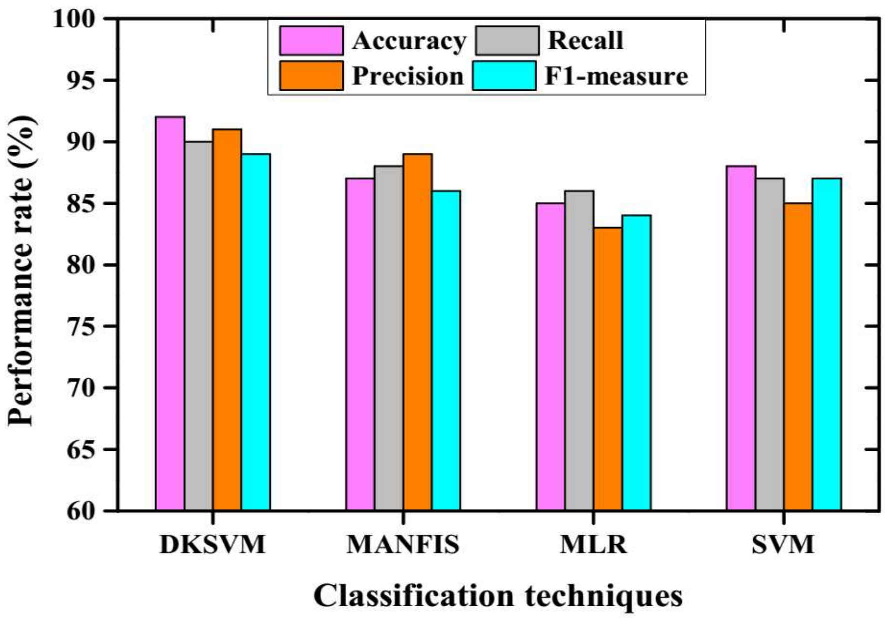

Figure 2. Comparative analysis using different performance metrics

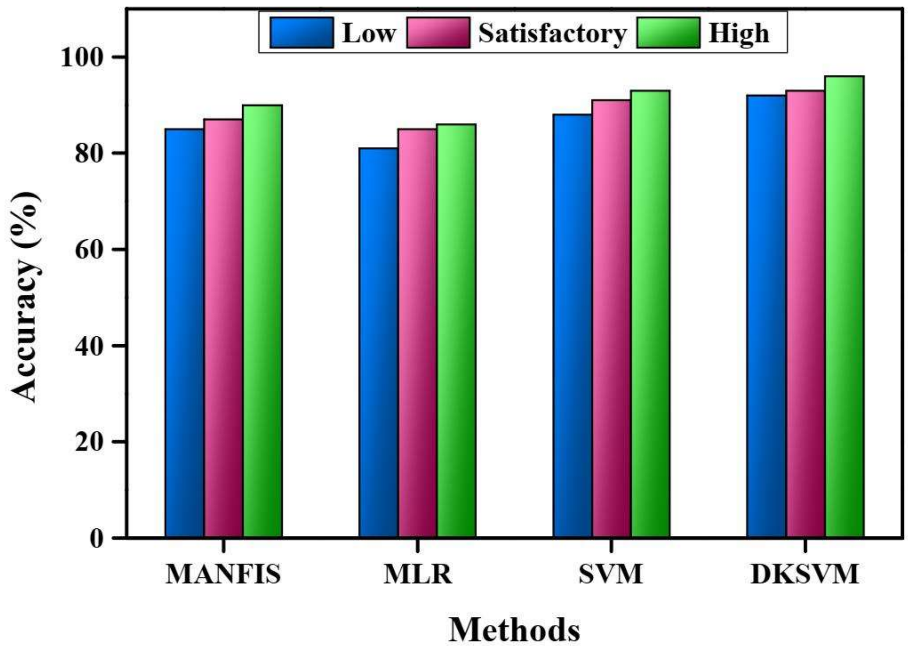

Figure 3. Comparison of accuracy for various performances

The accuracy is compared using the skillset of different batches of graduates and the results are demonstrated in Figure 3. The proposed approach is compared with various other existing approaches such as MANFIS, MLR, and SVM to determine three different performances such s low, satisfactory, and high. From the analysis, it is demonstrated that the proposed approach attained a high accuracy rate for all three classes.

The proposed model is compared with a few techniques like MANFIS, MLR, and SVM in terms of F1-score and test time, and the results are shown in Table 3. The proposed model offers an F1-score of 88%, 90%, and 92% for the low, satisfactory, and high classes.

Table 3. Performance evaluation using F-measure and test time

Techniques	F1-measure			Test time for every instance (ms)
	Low	Satisfactory	High	
Proposed DK-SVM	88	90	92	22.61
MANFIS (Ulloa, 2021)	87	88	89	23.57
MLR (Almufarreh, 2023)	84	85	86	25.23
SVM (Mengash, 2020)	86	87	90	24.12

5. CONCLUSION

The study proposes the Dual Kernel Support Vector Machine (DK-SVM) as a method to efficiently detect the skills of graduates. The DK-SVM utilizes two kernel functions, linear and radial basis, to improve accuracy. Data from King Abdulaziz University including admission scores and gender are collected. The model's performance is evaluated using random holdout and k-fold cross-validation methods with a split of 80% and 20% for training and testing respectively. Various performance metrics such as recall, precision, accuracy, and F1-score are employed for evaluating the model. The results show the proposed DK-SVM outperforms existing methods such as Support Vector Machine (SVM), Multiple Linear Regression (MLR), Multi-layer Adaptive Neuro-Fuzzy Inference System (MANFIS) in terms of precision, accuracy, recall, and F1-score, achieving an accuracy of 92%, a recall of 90%, a precision of 91%and an F1-score of 89%. Therefore, the proposed DK-SVM is an efficient method for predicting the skills of graduates. Meanwhile, this paper analyzed only education-related data and failed to analyze various other applications. Further research could explore the use of other feature selection and extraction methods in order to enhance the efficiency and accuracy of the DK-SVM. The proposed method will also be extended to identify the skills and strengths of individuals in a wider range of fields, beyond education. Additionally in the future, greater datasets and many diverse datasets could be used in this study in order to improve the generalizability of the results.

References

Abdelmoumin, G., Rawat, D.B. and Rahman, A., 2021. On the performance of machine learning models for anomaly-based intelligent intrusion detection systems for the internet of things. IEEE Internet of Things Journal, 9(6), pp.4280-4290.

Almufarreh, A., Noaman, K.M. and Saeed, M.N., 2023. Academic Teaching Quality Framework and Performance Evaluation Using Machine Learning. Applied Sciences, 13(5), p.3121.

Alam, T.M., Mushtaq, M., Shaukat, K., Hameed, I.A., Umer Sarwar, M. and Luo, S., 2021. A novel method for performance measurement of public educational institutions using machine learning models. Applied Sciences, 11(19), p.9296.

Alshanqiti, A. and Namoun, A., 2020. Predicting student performance and its influential factors using hybrid regression and multi-label classification. IEEE Access, 8, pp.203827-203844.

Alhazmi, E. and Sheneamer, A., 2023. Early Predicting of Students Performance in Higher Education. IEEE Access, 11, pp. 27579-27589.

McMurray, S., Dutton, M., McQuaid, R. and Richard, A., 2016. Employer demands from business graduates. Education+ Training.

James, S., Warhurst, C., Tholen, G. and Commander, J., 2013. What we know and what we need to know about graduate skills. Work, employment and society, 27(6), pp.952-963.

Mengash, H.A., 2020. Using data mining techniques to predict student performance to support decision making in university admission systems. IEEE Access, 8, pp.55462-55470.

Chui, K.T., Liu, R.W., Zhao, M. and De Pablos, P.O., 2020. Predicting students' performance with school and family tutoring using generative adversarial network-based deep support vector machines. IEEE Access, 8, pp.86745-86752.

Chui, K.T., Fung, D.C.L., Lytras, M.D. and Lam, T.M., 2020. Predicting at-risk university students in a virtual learning environment via a machine learning algorithm. Computers in Human Behavior, 107, p.105584.

Meegahapola, L.G. and Thilakarathne, C., 2019. Dynamic learner-assisted interactive learning tools for power systems engineering courses. IEEE Transactions on Education, 62(2), pp.149-156

Qiu, W., Tang, Q., Zhu, K., Yao, W., Ma, J. and Liu, Y., 2020. Cyber spoofing detection for grid distributed synchrophasor using dynamic dual-kernel svm. IEEE Transactions on Smart Grid, 12(3), pp.2732-2735

Rivas, Alberto, Jesús M. Fraile, Pablo Chamoso, Alfonso González-Briones, Sara Rodríguez, and Juan M. Corchado. "Students performance analysis based on machine learning techniques." In Learning Technology for Education Challenges: 8th International Workshop, LTEC 2019, Zamora, Spain, July 15–18, 2019, Proceedings 8, pp. 428-438. Springer International Publishing, 2019.

Ulloa-Cazarez, R.L., García-Díaz, N. and Soriano-Equigua, L., 2021. Multi-layer adaptive fuzzy inference system for predicting student performance in online higher education. IEEE Latin America Transactions, 19(01), pp.98-106.

Waheed, H., Hassan, S.U., Aljohani, N.R., Hardman, J., Alelyani, S. and Nawaz, R., 2020. Predicting academic performance of students from VLE big data using deep learning models. Computers in Human behavior, 104, p.106189.